Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Fusarium oxysporum trypsin at atomic resolution at 100 and 283 K: a study of ligand binding

Wojciech R. Rypniewski,[a]* Peter R. Østergaard,[b] Mads Nørregaard-Madsen,[b] Miroslawa Dauter[a]† and Keith S. Wilson[c]

[a]European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, [b]Novo Nordisk, Novo Alle, DK-2880 Bagsvaerd, Denmark, and [c]York University, Chemistry Department, Heslington, York YO1 5DD, England

† Present address: Brookhaven National Laboratory, Chemistry Department, Upton, NY 11973, USA.

Correspondence e-mail: wojtek@embl-hamburg.de

Received 21 July 2000
Accepted 9 October 2000

PDB References: TRY-N, 1gdu; TRY-F, 1gdq; TRY-ARG, 1fn8; TRY-LYS, 1gdn; TRY-LYS2, 1fy5; TRY-GLN, 1fy4.

The X-ray structure of *F. oxysporum* trypsin has been determined at atomic resolution, revealing electron density in the binding site which was interpreted as a peptide bound in the sites S1, S2 and S3. The structure, which was initially determined at 1.07 Å resolution and 283 K, has an Arg in the S1 specificity pocket. The study was extended to 0.81 Å resolution at 100 K using crystals soaked in Arg, Lys and Gln to study in greater detail the binding at the S1 site. The electron density in the binding site was compared between the different structures and analysed in terms of partially occupied and overlapping components of peptide, solvent water and possibly other chemical moieties. Arg-soaked crystals reveal a density more detailed but similar to the original structure, with the Arg side chain visible in the S1 pocket and residual peptide density in the S2 and S3 sites. The density in the active site is complex and not fully interpreted. Lys at high concentrations displaces Arg in the S1 pocket, while some main-chain density remains in sites S2 and S3. Gln has been shown not to bind. The free peptide in the S1–S3 sites binds in a similar way to the binding loop of BPTI or the inhibitory domain of the Alzheimer's β-protein precursor, with some differences in the S1 site.

1. Introduction

Trypsin (E.C. 3.4.21.4) is a serine proteinase characterized by specific cleavage of the peptide bond on the C-terminal side of lysine or arginine. Like other serine proteinases, it has been studied extensively in terms of its reaction mechanism, substrate specificity, inhibition and evolution (Czapinska & Otlewski, 1999; Neurath, 1984). The structural basis for the catalysis and specificity was established in a series of pioneering structural studies (Freer *et al.*, 1970; Henderson, 1970; Kraut *et al.*, 1972; Wright *et al.*, 1969) and serine proteinases have since proven a very fertile ground for studying structure–function relationships in enzymatic mechanism and intermolecular recognition. This is so because serine proteinases, while having a common reaction mechanism, show a very broad variability in substrate specificity, inhibition and biological function and are found to play key roles in diverse organisms. The reaction mechanism has been studied in detail (Blow, 1976; Kraut, 1977; Kossiakoff & Spencer, 1981; Kuhn *et al.*, 1998; Moult *et al.*, 1985), but the exact relationship between the catalytic residues and the sequence of events during the reaction has not been fully elucidated, especially for the catalytic serine. The accepted reaction mechanism proceeds *via* a tetrahedral transition-state intermediate. This is hard to observe owing to its transient nature (West *et al.*, 1990), but a number of model compounds

have been found to give rise to tetrahedral covalently bound adducts at the serine (Chen *et al.*, 1995; Mac Sweeney *et al.*, 2000; Rypniewski, Dambmann *et al.*, 1995). Several authors have also reported enzyme–product complexes (Harel *et al.*, 1991), acyl-enzyme intermediates (Strynadka *et al.*, 1992) or even a trapped tetrahedral reaction intermediate (Yennawar *et al.*, 1994). However, working with natural substrates is difficult because of the dynamic nature of their interactions with the enzymes. In the case of proteinases, an additional constraint is the enzymes' susceptibility to autolysis. For this reason, investigation of substrate specificity and catalysis has relied mainly on enzyme complexes with inhibitors or substrate analogues. The binding site of trypsin and related enzymes has been characterized using naturally occurring inhibitors (Bode & Huber, 1992) interacting with the enzyme in a substrate-like manner, according to the 'standard mechanism' (Laskowski & Kato, 1980), and their complexes with the enzymes have been determined in detail (Finer-Moore *et al.*, 1992; Scheidig *et al.*, 1997). The characterized binding site spans subsites S3–S3′ for the corresponding amino acids P3–P3′ of the peptide ligand notation by Schechter (Schechter & Berger, 1967), with hydrolysis taking place between P1 and P1′. The subsite S1 also contains the 'specificity pocket', which in trypsin binds selectively basic side chains. The structural and energetic aspects of the binding of the main chain and in the S1 pocket have been probed in detail (Helland *et al.*, 1999; Lu, Apostol *et al.*, 1997; Lu, Qasim *et al.*, 1997) and reviewed recently (Czapinska & Otlewski, 1999).

The trypsin from *F. oxysporum* has been investigated previously, inhibited by diisopropylfluorophosphate (DFP), which forms an irreversible covalently bound adduct at the catalytic serine (Rypniewski, Dambmann *et al.*, 1995). However, the enzyme is also stable for extensive periods of time in the presence of boric acid, which acts as a mild reversible inhibitor. We have found that the protein can be crystallized under these conditions and the crystals give diffraction of outstanding quality. In the early stages of this investigation we realised that the binding site is occupied by its own autolysis products. This led to further studies in an attempt to use the power of synchrotron radiation on crystals diffracting to ultrahigh resolution to observe the details of weakly binding ligands.

The present paper presents results of X-ray studies of the *F. oxysporum* trypsin with its own autolysis products and in the presence of the added amino acids arginine, lysine and glutamate, in order to investigate their binding in the S1 specificity pocket. The advantages of very high resolution become clear, but the final interpretation of low-occupancy alternative chemical structures still continues to stretch the method to its limits.

The following structures are presented.

(i) TRY-N (PDB code 1gdu): native structure of *F. oxysporum* trypsin at 283 K at 1.07 Å resolution.

(ii) TRY-F (1gdq): the native structure at 100 K at 0.97 Å resolution.

(iii) TRY-ARG (1fn8): the trypsin structure in the presence of 20 m*M* arginine at 0.81 Å.

(iv) TRY-LYS (1gdn): as above, soaked in 20 m*M* lysine, at 0.81 Å.

(v) TRY-LYS2 (1fy5): as above, in the presence of 170 m*M* lysine, at 0.81 Å.

(vi) TRY-GLN (1fy4): as above in the presence of 20 m*M* glutamine, at 0.81 Å.

2. Materials and methods

2.1. Crystallization

The protein was prepared as described previously (Rypniewski *et al.*, 1993). The protein solution used for crystallization initially contained 10 mg ml^{-1} *F. oxysporum* trypsin, 100 m*M* boric acid, a reversible inhibitor used to protect the protein against autolysis, 2 m*M* $CaCl_2$, 100 m*M* NaCl and 10 m*M* 3,3-dimethyl glutaric acid (Sigma) pH 6.0. The protein was concentrated to 25 mg ml^{-1} by centrifugation through Centricon filters. Crystals were grown by sitting-drop vapour diffusion. The wells contained 1.4 *M* Na_2SO_4, 0.1 *M* sodium citrate pH 5.0–6.0 and 0–5% 2-propanol. The droplets were prepared by mixing equal volumes of the above protein solution and the solution from the well. Although the crystallization conditions were similar to those used to obtain crystals of DFP (diisopropylfluorophosphate) inhibited protein (Rypniewski, Mangani *et al.*, 1995), no crystals could initially be grown in the absence of DFP and once they were obtained they turned out to be triclinic rather than the DFP-inhibited monoclinic form. The first crystals were obtained only after 5% 2-propanol was included in an attempt to mimic closely the conditions for the inhibited crystals. Subsequently, DFP-free crystals were easily grown by microseeding with or without 2-propanol.

2.2. Data collection and processing

All data sets were measured from single crystals and diffraction data were collected using synchrotron radiation on the EMBL beamlines at the DORIS storage ring, DESY, Hamburg, with a MAR Research imaging-plate scanner, except for part of the low-resolution scan for TRY-N which, owing to beamtime constraints, was collected on a sealed-tube X-ray generator with a molybdenum target. The native crystal used for data collection at ambient temperature was mounted in a quartz capillary, which was then placed on the goniostat in a stream of chilled air that was used to maintain a temperature of ∼283 K at the crystal during the data collection. The native crystal used for cryogenic data collection was mounted in a cryoloop after a brief immersion in solution containing 20% glycerol in addition to the ingredients present in the crystallization well. Leaving the crystal in such a cryoprotectant for longer than ∼10 s caused the crystals to crack, whereas using other cryoprotectants such as 2,5-methylpentanediol or ethylene glycol in concentrations sufficient to prevent severe ice formation resulted in salt crystals appearing within seconds. Therefore, the crystals used in ligand-binding studies were prepared by soaking in solutions containing increasing amounts of glycerol (0, 5, 10, 15, 20%) in addition to ingre-

dients similar to those in the crystallization well, buffered at pH 6, and the following concentrations of amino acids buffered at pH 6: 20 m*M* arginine, 20 m*M* lysine, 170 m*M* lysine and 20 m*M* glutamine for structures TRY-ARG, TRY-LYS, TRY-LYS2 and TRY-GLN, respectively. In order to inhibit formation of salt crystals, each solution was Millipore filtered and centrifuged prior to the experiment; glassware was siliconized and rinsed under distilled water to remove dust particles which might act as nucleation sites. The protein crystals were transferred using cryoloops freshly rinsed with distilled water. The crystals were observed under a dissecting microscope: freshly transferred crystals were buoyant, but after ∼2 min they sank into the solution. They were then transferred to the next solution. Despite these precautions, some crystals developed cracks and were discarded.

A range of reciprocal space of at least 180° was covered in several sweeps at different exposure times to record the full dynamic range of intensities. The programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) were used for data reduction and scaling. Initial scaling included refinement of the exponential factor to check for resolution-dependent decrease of intensity during data collection. No significant decay was observed and subsequently the images were scaled without a relative temperature factor. The intensities were converted to structure-factor amplitudes and a correction was applied to weak or negative measurements (French & Wilson, 1978); however, for the refinement with *SHELXL* the original intensities were used.

2.3. Structure determination

The atomic resolution structure of the *F. oxysporum* trypsin at 1.07 Å was determined by molecular replacement as implemented in the program *AMoRe* (Navaza, 1994) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The rotation function was calculated using terms between 10 and 3 Å, with a Patterson search radius of 20 Å. A solution was obtained using the coordinates of the DFP-inhibited *F. oxysporum* trypsin previously determined at 1.55 Å resolution (PDB code 1try; Rypniewski, Dambmann *et al.*, 1995), placed in a *P*1 cell of dimensions 70 × 70 × 70 Å.

2.4. Refinement

The models were initially refined against all the diffraction data using *REFMAC* (Collaborative Computational Project, Number 4, 1994) with isotropic atomic temperature factors and including contributions to the structure factors of the H atoms that could be defined in standard geometry. Subsequently, the models were refined against diffraction intensities rather than structure-factor amplitudes using *SHELXL* (Sheldrick & Schneider, 1997) by the conjugate-gradient algorithm including anisotropic atomic temperature factors and H atoms, generated either in their riding positions or, in the case of mobile H atoms, by the electron-density search algorithm incorporated in *SHELXL*. Water molecules were modelled using *ARP* with default settings (Lamzin & Wilson, 1993). The significance level for defining new waters in the $(F_o - F_c)$ map was determined automatically; it started at 3σ and approached the 4σ level at the end of the refinement. Water molecules were removed from the model if they were in density less than 0.5σ in the $(3F_o - 2F_c)$ map. The models were inspected against $(3F_o - 2F_c)$ and $(F_o - F_c)$ maps and adjusted manually between cycles of refinement using *TURBO-FRODO* (Roussel & Cambillau, 1991) running on a Silicon Graphics O2 work station.

In the later stages of the refinement occupancy factors were refined for water molecules. In addition to the above rejection criteria, water molecules were removed from the model if their occupancy factor dropped below 0.5. Occupancy was also refined for the ligands bound in the active site. In the structures solved in the presence of added amino acids occupancy factors were refined separately for the S1 site from the S2–S3 sites, allowing for the possibility that the added amino acids may target the specificity pocket. In the native structures the ligand bound in S1–S3 sites was refined with a single occupancy factor.

Occupancy factors were also refined for groups that showed a loss of intensity in the $(3F_o - 2F_c)$ and $(F_o - F_c)$ maps, except where the apparent lack of intensity was likely to be a consequence of disorder, as in the case of surface side chains. The refined groups include all the thiols of cysteines and the hydroxyl group of the catalytic Ser195 in all the structures except TRY-N, which showed good density. In TRY-LYS2, which showed a visible loss of density for several other groups, especially carboxyl and hydroxyl, the occupancy of these were also refined as well as the occupancy of all the other carboxyl groups and the hydroxyl groups of serines, tyrosines and threonines, except for disordered side chains, and the thiomethyl group of Met180.

The ligands in the active site were identified by initially filling the electron density with free O atoms (water). Modelling of the bound peptides was aided by comparing the

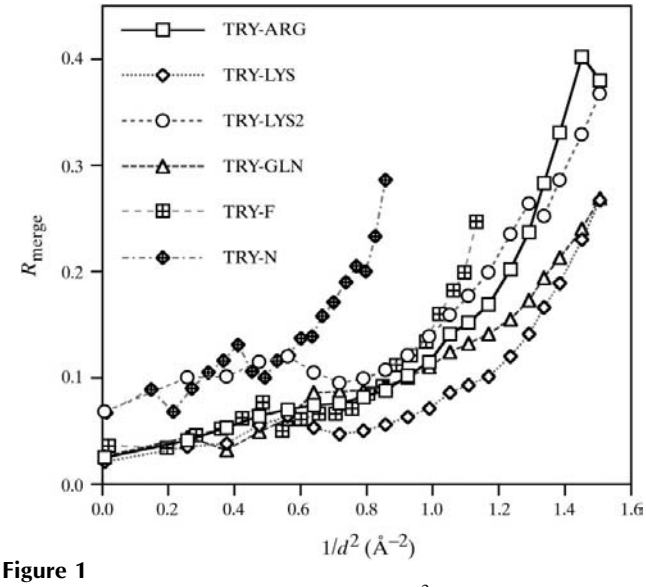


Figure 1
R_{merge} $(\sum|I_i - \langle I\rangle|/\sum\langle I\rangle)$ as a function of $1/d^2$, where d is the resolution defined in Å.

Table 1
Summary of data.

	TRY-N	TRY-F	TRY-ARG	TRY-LYS	TRY-LYS2	TRY-GLN
X-ray source	BW7B/X31/Mo	BW7B	BW7B	BW7B	BW7B	BW7B
Temperature (K)	283	100	100	100	100	100
Resolution range (Å)	15–1.07	12–0.93	20–0.81	20–0.81	20–0.81	20–0.81
Wavelength (Å)	0.950/0.920/0.71	0.870	0.8468	0.8468	0.8443	0.8468
Number of images	272	361	547	457	582	663
Exposure time/high-resolution image (s)	∼10	∼30	∼60	∼90	∼120	∼120
Oscillation range (°)	1.5–4.0	1.0–3.0	1–2.5	1–2.5	1–2.5	1–2.5
R_{merge}†	0.083	0.043	0.037	0.030	0.082	0.038
Mosaicity (°)	0.46	0.75	0.43	0.40	0.57	0.35
Raw measurements used	297606	576525	990347	700099	1208333	1184962
Unique reflections	70387	107781	158550	156398	165439	159467
Completeness (%)	92.8	92.4	92.6	90.9	96.4	92.8
Completeness in high-resolution bin (%)	87.6	86.3	82.5	81.9	85.0	88.4
Greater than 3σ (%)	56.9	84.1	73.1	76.5	80.4	90.5
$I/\sigma(I)$ in high-resolution bin	1.4	2.3	1.5	1.8	2.2	3.1
Unit-cell parameters						
a (Å)	33.29	33.05	33.15	33.16	33.18	33.13
b (Å)	37.05	36.97	36.70	36.76	36.78	36.74
c (Å)	40.26	39.78	39.56	39.62	39.60	39.64
α (°)	102.46	102.81	102.09	102.19	102.18	102.24
β (°)	103.81	104.56	104.75	104.70	104.50	104.75
γ (°)	102.07	102.40	103.01	102d91	103.12	102.85

† $R_{merge} = \sum|I_i - \langle I\rangle|/\sum\langle I\rangle$, where I_i is an individual intensity measurement and $\langle I\rangle$ is the average intensity for this reflection, with summation over all data.

electron density of the different structure and using as a guide complexes of bovine trypsin with BPTI (PDB code 2ptc; Marquart *et al.*, 1983) and APPI inhibitor (PDB code 1taw; Scheidig *et al.*, 1997). After the peptides were included in the model, the remaining density was modelled as water molecules.

The refinement was terminated when no further significant improvement could be achieved in *R*-factor statistics, model completeness and stereochemistry and electron density. After the refinement was complete, one additional cycle was run using full-matrix least-squares minimization to estimate from the correlation matrix the standard deviations for atomic positions and temperature factors. For reasons of limited program memory, the structure was divided into blocks of approximately 20 residues with a one-residue overlap between the blocks. All restraints and damping were removed in this cycle. Standard deviations for distances were derived from standard deviations for the atomic positions. Standard errors were calculated for mean values.

3. Results

3.1. Quality of the data and the final models

The data collection and final statistics are summarized in Table 1. Fig. 1 shows R_{merge} as a function of the resolution for all the data sets. TRY-N and TRY-LYS2 have poorer merging statistics at low resolution than the other data sets. For TRY-N this is probably a consequence in part of the low-resolution data having been collected using a sealed-tube X-ray generator. Attempts were made to recollect room-temperature data using only synchrotron radiation but there was always significant intensity decay at high resolution. In the end, despite having been obtained partly from a sealed-tube source, the TRY-N data set was deemed the most acceptable because it showed no radiation damage at high resolution. For TRY-LYS2 the likely cause for the relatively poor merging statistics at low resolution is the excessive speed of the φ-motor in the lowest resolution scan, despite using the 'time' mode data collection, which allows faster scan speeds than the 'dose' mode. Subsequent attempts to collect data from crystals soaked in a high concentration of lysine were unsuccessful owing to crystal cracking. In the end the TRY-LYS2 data were deemed acceptable: despite showing higher than average R_{merge} at low resolution this data set has the highest redundancy and completeness.

The unit-cell parameters decrease by up to 1% between the ambient-temperature and the cryogenic structures, with the liganded crystals, which were extensively soaked in the cryoprotectant, having the smallest unit cell, while the native cryo

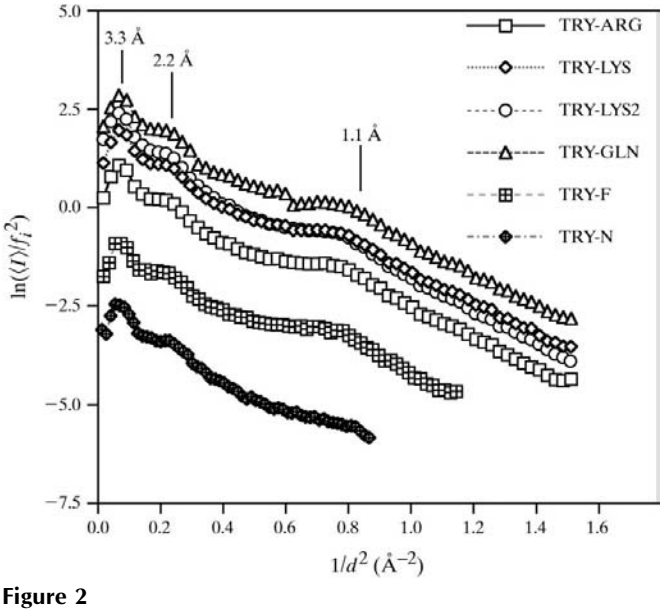


Figure 2
Wilson plots for the data. The differences in the ordinate for the different data sets reflect approximately the differences in the recorded intensity of data. Characteristic humps can be seen at approximately 1.1, 2.2 and 3.3 Å owing to the non-random distribution of the atoms.

Table 2
Summary of the models and estimated standard deviations (e.s.d.s) in coordinates.

	TRY-N	TRY-F	TRY-ARG	TRY-LYS	TRY-LYS2	TRY-GLN
No. of non-H atom positions†	1975	2243	2220	2329	2130	2319
Protein	1610	1704	1792	1868	1738	1905
Water molecules	360	523	401	432	363	385
Other atoms	5 (1 sul.)	16 (2 sul., 1 glyc.)	29 (1 sul., 4 glyc.)	29 (1 sul., 4 glyc.)	29 (1 sul., 4 glyc.)	29 (1 sul., 4 glyc.)
Temperature factors						
Overall	18.0	13.5	13.0	12.9	13.6	12.9
Protein	13.6	9.3	9.9	9.9	10.9	10.3
Main chain	11.8	8.1	8.7	8.8	9.4	9.2
Side chain	15.8	10.5	10.9	10.7	12.2	11.0
Solvent	37.6	26.8	26.5	25.3	26.2	25.3
R factor‡	0.1036	0.0986	0.1085	0.1083	0.1242	0.1080
E.s.d.s§						
All 'heavy' atoms	0.067	0.070	0.051	0.057	0.054	0.052
Protein	0.046	0.045	0.038	0.041	0.040	0.040
Protein O	0.038	0.043	0.031	0.033	0.035	0.033
Protein N	0.034	0.046	0.039	0.039	0.040	0.041
Protein C	0.051	0.046	0.041	0.043	0.041	0.041
Protein S	0.007	0.028	0.034	0.040	0.048	0.041
Main chain	0.029	0.038	0.028	0.029	0.028	0.028
Main-chain O	0.024	0.033	0.025	0.027	0.026	0.028
Main-chain N	0.025	0.037	0.026	0.025	0.025	0.025
Main-chain C	0.033	0.040	0.031	0.031	0.030	0.030
Solvent	0.133	0.122	0.089	0.103	0.097	0.093

† The number of atomic positions, unlike the number of atoms, takes into account multiple conformations and reflects approximately the number of independent variables. There are 1570 non-H atoms in *F. oxysporum* trypsin and approximately 1450 H atoms. ‡ $R = \sum|F_o - F_c|/\sum F_o$. § All e.s.d. statistics exclude H atoms, which were not refined.

structure (TRY-F), which was only briefly exposed to the cryoprotectant, has intermediate unit-cell parameters and the highest mosaicity.

Fig. 2 shows the Wilson plots for all data (Wilson, 1942). Fitting a straight line to the intensity distribution gives an estimation of the overall *B* factor (6–8 Å^2) which is quite different from the average temperature factor for the atomic models (Table 2). The difference is mainly a consequence of the broad 'hump' in the intensity distribution centred around 1.1 Å. This and similar but smaller features at 2.2 and 3.3 Å arise from non-random distribution of atoms, which results in non-zero off-diagonal terms in the summation over $\mathbf{F(S)}\cdot\mathbf{F}^*\mathbf{(S)}$ (Murshudov, personal communication). If data beyond 1.8 Å are omitted, the obtained estimates of the *B* factor are much closer to the values derived from the atomic models.

Table 3 summarizes the stereochemical restraints used in the refinement and the final deviations from the target values. Table 2 gives a summary of the refined models and the estimated standard uncertainties (e.s.u.s). The e.s.u.s in atomic positions generally reflect the electron number of the atoms, with discernible differences between oxygen, nitrogen and carbon. The uncertainties are generally lower for main-chain protein atoms and much higher for solvent atoms and poorly resolved side chains. The Ramachandran plots (Ramachandran *et al.*, 1963) for the structures have 90% of the residues lying in the most favoured regions while 10% fall within the allowed regions as defined in the program *PROCHECK* (Laskowski *et al.*, 1993). The other parameters analysed by *PROCHECK* were close to the average values for atomic resolution structures.

3.2. The overall structure

All the models show the characteristic trypsin fold consisting of two closely associated domains (Fig. 3); the core of each is a six-stranded antiparallel β-barrel with the first four strands forming a Greek key motif. The N-terminus is embedded in the C-terminal domain, while the C-terminus forms an α-helix reaching the N-terminal domain. The catalytic triad of residues, Ser195, His57 and Asp102, are located in the cleft between the two domains, while the specificity pocket that determines the substrate specificity in the trypsin family of serine proteinases is located in the C-terminal domain with its entrance close to the catalytic site at the Ser195. The secondary structure of *F. oxysporum* trypsin is essentially identical to that described previously for the monoclinic crystal structure and will not be described here (Rypniewski, Dambmann *et al.*, 1995). No bound metal ions were observed based on density and coor-

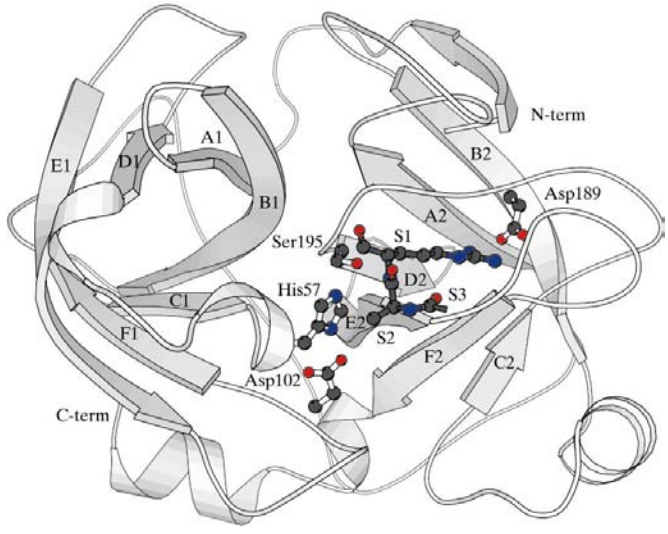


Figure 3
Schematic representation of TRY-N. Secondary-structure elements forming the N-terminal domain are marked A1–F1 and the C-terminal domain A2–F2 as defined previously (Rypniewski, Mangani *et al.*, 1995). The active site is indicated by the side chains of the 'catalytic triad': Ser195, His57 and Asp102. The side chain of Asp189 is also shown at the bottom of the 'specificity pocket' defining the tryptic substrate specificity. The peptide bound in sites S1 to S3 is indicated by a ball-and-stick model with dark bonds.

Table 3
Weighting parameters for the least-squares refinement and the final deviations from target values.

		Deviation from target values (number of parameters)					
Distances (Å)	Weight†	TRY-N	TRY-F	TRY-ARG	TRY-LYS	TRY-LYS2	TRY-GLN
Bond lengths (1–2 neighbours)	0.020	0.015 (1637)	0.014 (1732)	0.015 (1825)	0.016 (1905)	0.016 (1771)	0.016 (1943)
Bond angles (1–3 neighbours)	0.040	0.031 (2230)	0.031 (2361)	0.035 (2498)	0.036 (2609)	0.034 (2420)	0.037 (2662)
Chiral volumes	0.100	0.102 (264)	0.112 (277)	0.095 (294)	0.134 (308)	0.134 (284)	0.132 (314)

† The weights correspond to $1/\sigma^2$.

Table 4
Refined occupancy factors of ligands in the binding site.

	TRY-N	TRY-F	TRY-ARG	TRY-LYS	TRY-LYS2	TRY-GLN
P1	0.51	0.33	0.43	0.39	0.59	0.25
P2–P3		—	0.31	0.33	0.42	0.17
Water, or possibly other atoms, at S1 active site						
2003	0.29	0.38	0.36	0.37	0.28	0.33
2004	0.69	0.67	0.46	0.47	—	0.54
2005	0.59	0.38	0.34	0.32	—	0.31
2006	0.45	0.36	0.22	0.35	—	0.40
Waters in S1 specificity pocket						
2101	0.55	0.60	0.56	0.60	0.46	0.61
2102	0.47	0.70	0.54	0.83	0.58	0.66
2103	0.44	0.79	0.64	0.55	0.60	0.37
2104	0.88	0.77	0.66	0.70	0.60	0.73
2105	0.89	1.00	0.98	1.00	1.00	1.00
Waters at S2–S3 site						
2201	0.53	0.77	0.63	0.62	0.54	0.71
2202	—	0.85	0.58	0.71	0.85	0.62
2203	—	0.87	—	—	—	—
2204	—	0.28	0.41	0.56	0.34	0.43

dination criteria, confirming the earlier finding that unlike other trypsins *F. oxysporum* trypsin contains no Ca^{2+}-binding sites. This is significant in view of a suggestion that the trypsin-like serine proteases' requirement for calcium is a safety feature against an accidental activation in the cytoplasm, where the calcium concentration is low (Kretsinger, 1976).

The models presented here all have high overall similarity. Most similar are the structures from crystals soaked with amino acids, TRY-ARG, TRY-LYS, TRY-LYS2 and TRY-GLN (long stepwise soak in cryoprotectant), with r.m.s. differences in C^{α} atoms between least-squares fitted models of approximately 0.05 Å, counting only the major conformation in the case of main-chain disorder. The difference between the amino-acid soaked structures and the native cryo structure, TRY-F (rapid soak in cryoprotectant), is 0.2–0.3 Å; with the native model, TRY-N, at ambient temperature it is 0.2 Å. The difference between the two native models is also 0.2 Å. The difference between any of the structures and *F. oxysporum* trypsin solved from monoclinic crystals (Rypniewski, Dambmann *et al.*, 1995) is approximately 0.3 Å.

The r.m.s. difference in C^{α} atoms between TRY-N and the bovine β-trypsin (PDB code 5ptp; Finer-Moore *et al.*, 1992) is 1.4 Å using 210 atoms after excluding C^{α} pairs deviating by more than 3 r.m.s. With the trypsin from *S. griseus* (PDB code 1sgt; Read & James, 1988), the difference is also 1.4 Å using 195 C^{α} atoms selected likewise. The differences are 0.6 Å with both proteins if only the 58 conserved core of residues (Rypniewski, Dambmann *et al.*, 1995) are compared.

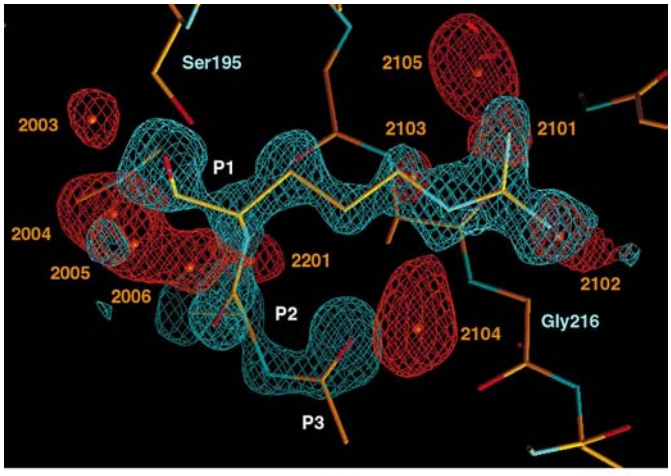
3.3. Ligand binding

3.3.1. TRY-N and TRY-F. When the TRY-N electron density was initially examined it revealed density closely resembling an Arg side chain filling the S1 specificity pocket (Fig. 4). There was also density close to the hydroxyl group of the catalytic Ser195 and in the S2 site reaching to S3. A comparison with the bovine trypsin–BPTI complex (PDB code 2ptc; Marquart *et al.*, 1983) served as an initial guide in building a short peptide into the density. No significant density was observed to trace the chain further upstream and similarly there was no density for the P1′ residue or residues further downstream.

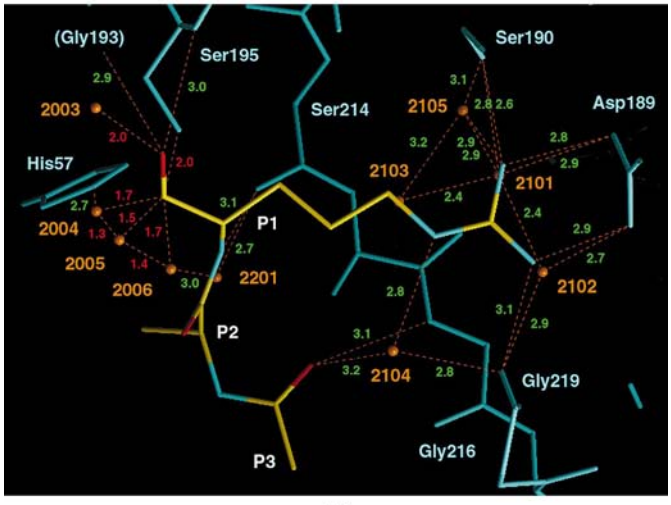
Examination of TRY-F revealed a substantial reduction of density in the active site compared with TRY-N. This was interpreted as the result of displacement of the peptide during the exposure of the crystal to the cryoprotectant buffer. Several strong peaks remained in the specificity pocket and were interpreted as water molecules based on their proximity to hydrogen-bonding groups and to one another. Weak density remained for the Arg side chain and was included in the model. The peak for C^{α} was very weak; N was not observed. No density was observed for residues P2 and P3, but there was density for several water molecules in the sites S2 and S3. Several peaks were observed near Ser195 but their interpretation was difficult. Their density was probably an average of alternatively bound partially occupied solvent water, residual borate, which was present in the enzyme buffer, and acyl-enzyme intermediate.

3.3.2. TRY-ARG. TRY-ARG revealed the binding site at higher resolution (Fig. 5*a*). There were small but distinct peaks for atomic positions of a peptide bound in sites S1, S2 and part of S3. The peptide appears continuous with normal stereochemistry between P1 and P2, although the S1 site probably contains a mixture of the originally bound peptide and the added Arg. The Arg density in the S1 site had lower occupancy than TRY-N but higher occupancy than TRY-F (Table 4). The occupancy for the P1 residue was refined independently from those for P2–P3 to allow for the introduction of the individual Arg at S1. Only C^{β} is identifiable of the P2 side chain. Subsequently, P2 was modelled as Ala. The carboxyl group of residue P3 has also been modelled. No density was observed for any peptide on the prime side of the binding site.

Arg binds with the side chain in the S1 pocket in an extended conformation similar to the binding found in inhibitor studies of chymotrypsin and trypsin (PDB code 1taw; Scheidig *et al.*, 1997). The guanidinium group forms a cyclic salt bridge with the carboxyl group of Asp189 (Fig. 5*b*). In addition, $N^{\eta1}$ forms a hydrogen bond with Ser190 OH and water 2105; $N^{\eta2}$ hydrogen bonds the carbonyl O atom of Gly219. N^{ε} forms a hydrogen bond with water 2104. The $N^{\eta1}$ and $N^{\eta2}$ sites overlap closely with the sites of two water molecules 2101 and 2102 which bind in the absence of the amino acid. The density of a third water molecule 2103 overlaps with C^{δ}. This region of the S1 pocket is hydrophobic and the water forms a short hydrogen bond only with water 2101 near the $N^{\eta1}$ site.



(*a*)



(*b*)

Figure 4
Ligand binding in TRY-N. (*a*) Two 'omit' ($F_o - F_c$) maps are shown: for the peptide model P1–P3 (blue contours) and solvent or other atoms (red). The maps were calculated after omitting peptide or water molecules from the refined model and ten cycles of least-squares minimization. For the omit map of the peptide, the occupancy factors of the water molecules overlapping the peptide were fixed at the previously refined values. The maps are contoured at 3σ. Water molecules, or possibly other atoms, near the catalytic site are labelled 2003–2006. Waters in the S1 specificity pocket are labelled 2101–2105. The water molecule in S2 site is labelled 2201. The catalytic Ser195 is shown together with Gly216 which forms part of the specificity pocket. (*b*) Geometry of the binding site. The bound peptide (P1–P3) is shown in colours, protein in blue, solvent and other sites in orange. Hydrogen-bonding distances are shown in green; distances between sites less than 2.0 Å apart are shown in red, corresponding to alternatively occupied sites or bonded atoms.

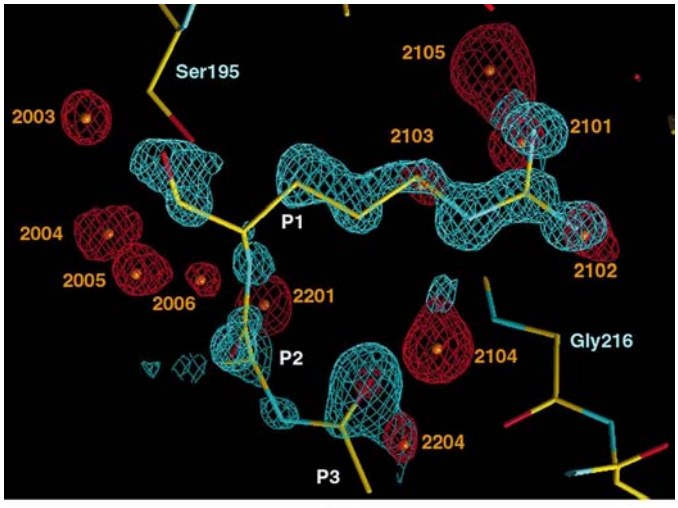
The water molecules which bind alternatively with the peptide were identified from the residual density after the bound peptide had been taken into account by modelling and refinement and by comparing the electron density for the different structures, especially TRY-GLN and TRY-F, where the peptide density is low. Apart from the three overlapping waters (2101, 2102 and 2103), two other water molecules (2104 and 2105) interact with the side chain in the specificity pocket at normal hydrogen-bonding distances, but the elongated shape of their density suggests closely located alternative positions, probably indicating small displacements of the waters as the ligand side chain enters the specificity pocket. Several water molecules were also identified close to residues P2 and P3 or overlapping with them as alternative positions. The peaks for waters 2101–2105 in the specificity pocket and 2201 in the S2 site are observed in all the structures. 2202 and 2204 are observed in all the 100 K structures. Table 4 summarizes the occupancy factors of the ligands in the binding site for the different models.

There were several peaks near the catalytic Ser195. Some of them were close to positions found or deduced in studies of acyl-enzyme intermediates or inhibitor complexes (Harel *et al.*, 1991; Saxena *et al.*, 1996; Yennawar *et al.*, 1994). The peak nearest the catalytic Ser was close to where the carbonyl C atom of the scissile peptide bond is expected to attach to the nucleophilic hydroxyl group. Another peak corresponded to the expected position of the carbonyl O atom bound in the 'oxyanion hole'. The peaks were interpreted accordingly. The close relative proximity of the remaining peaks in the catalytic site excludes the possibility that they are water sites occupied simultaneously. However, the sum of the occupancy factors for the neighbouring atoms does not significantly exceed unity, so the electron density could represent the superposition of alternative sites *e.g.* for water 2003 whose apparent location from the P1 carbonyl O atom is only 1.8 Å in TRY-ARG, 1.4 Å in TRY-F and 2.0 Å in TRY-N. Site 2003 is near the expected position of the 'Henderson water' – the disputed nucleophile in deacylation (Henderson, 1970). Water 2201 seems equivalent to the proposed alternative ('water 1082') to the 'Henderson water' (Singer *et al.*, 1993). The density of water 2201 overlaps with P1 and P2: the distance to P1 N is 1.2 Å, to P2 C 1.1 Å and to P2 C^{α} 1.4 Å – too close for the sites to be occupied simultaneously, as has been pointed out by Perona *et al.* (1993).

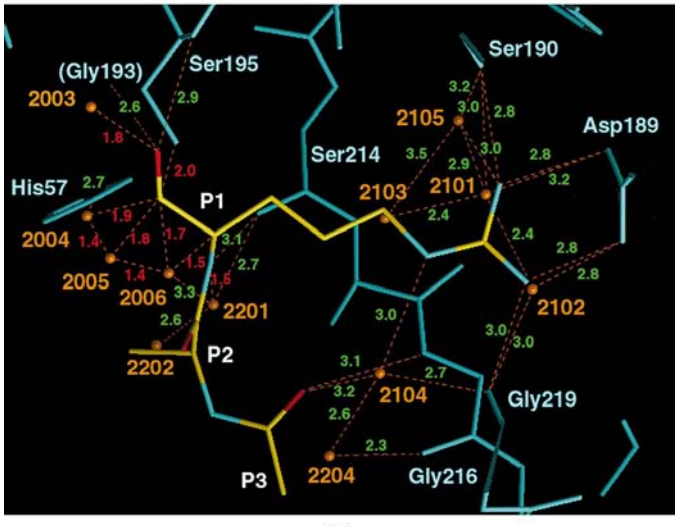
The remaining peaks, 2004, 2005 and 2006, are on the bulk-solvent side of the molecule. Their presence is independent of the presence of the other peaks in the active site (compare with Fig. 5 and TRY-LYS2 below), suggesting that despite close distances to P1 atoms they belong to a different molecular species. Apart from the possibility that they are alternative water sites, the density could in part belong to borate,

which could bind in the active site in a tetrahedral configuration possibly coinciding with the sites of P1 C, O, C^{α} and 2005. The presence of borate, which was included in the original protein preparation, cannot be proved or disproved based on the examination and comparison of the different structures.

3.3.3. TRY-LYS and TRY-LYS2. TRY-LYS revealed only weak density in the S1 pocket which could be interpreted as a Lys side chain but could also be partly a consequence of residual Arg. However, in TRY-LYS2 (Fig. 6) the Lys side-chain density is much stronger, although the definition of the side-chain atoms is poorer than in TRY-ARG (Fig. 5). The binding in the S1 pocket is similar to that observed in the BPTI–trypsin complex (PDB code 2ptc; Marquart *et al.*, 1983), with the Lys side chain in an extended conformation. The shorter side chain makes only rather long interactions with the bottom of the S1 pocket, with distances of 3.4 Å between N^{ζ} and each of the Asp189 carboxyl O atoms. It also interacts with Ser190 O (3.2 Å) and with water 2102. The centre of the peak is only 1.9 Å from N^{ζ} – too close for a normal hydrogen-bonded interaction – but the elongated density was modelled as two alternative locations. One refined at a hydrogen-bonding distance of 2.4 Å and the other as the alternative position when the side chain was not present. There is a weak indication from the 'omit' maps of two alternative positions also for the Lys side chain. The other position would put

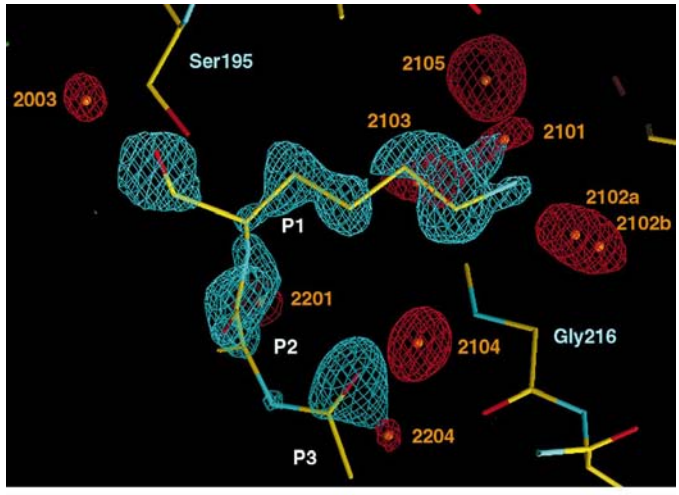


(*a*)

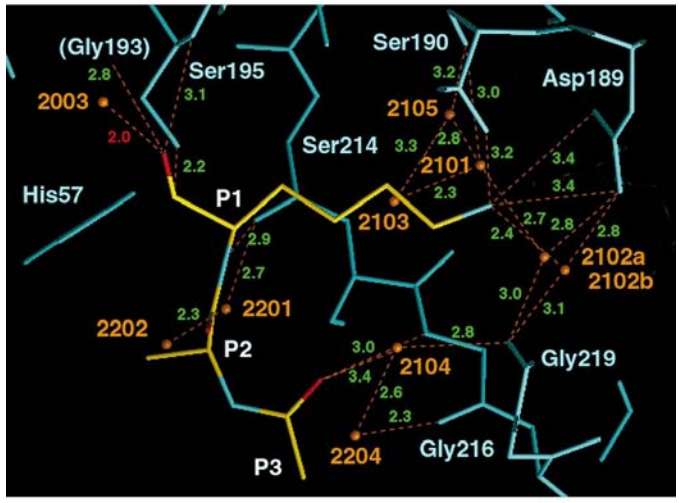


(*b*)

Figure 5
Ligand binding in TRY-ARG. (*a*). Omit maps calculated, colour coded and labelled as in Fig. 4. The peptide density is contoured at the 3σ level and the solvent density at 4σ. The peptide density has lower occupancy than for TRY-N (Fig. 4, Table 5) but its definition is higher. (*b*) Geometry of the binding site. The labelling scheme is similar to Fig. 4(*b*). Some atoms have been omitted for clarity.



(*a*)



(*b*)

Figure 6
Ligand binding in TRY-LYS2. (*a*) 'Omit' maps calculated as for Fig. 4(*a*), at the same contour levels and with similar colouring scheme. Sites 2004–2006 are not occupied in this structure, as shown by the absence of difference density. There is an indication of a double conformation of the Lys side chain, not modelled because of weak density level for refinement. The water at 2102 has been modelled in two alternative conformations: 2102b is at a bonding distance to Lys N^{ζ}, whilst 2102a is at clashing distance, binding in the absence of the peptide ligand. (*b*) Geometry of the binding site. The labelling scheme is similar to Fig. 4(*b*). Some atoms have been omitted for clarity.

N^{ζ} within hydrogen-bonding interaction distance with Ser190 OH. However, the density of the Lys side chain was deemed inadequate for the refinement of two different conformations and an average position was sought instead, although two alternative sites cannot be ruled out. Two other partially occupied water molecules overlap the Lys density. Their position is similar to the waters in TRY-ARG and the other structures.

There is weak but interpretable density for C^{α} and N of P1. Despite the presence of free Lys which has partly penetrated S1, there is still density for the peptide bound at sites S2 and part of S3, but density at the catalytic Ser195 is much reduced. The position for the P1 carboxyl group is poorly defined. The peak at 2003 ('Henderson water') is present at 2 Å from P1 O, but peaks 2004–2006 observed in TRY-ARG and TRY-F and TRY-N are absent. The density of the hydroxyl O atom of Ser195 is also significantly reduced in TRY-LYS2, with a refined occupancy factor of only 0.6.

3.3.4. TRY-GLN. Gln-soaked crystals revealed no density for Gln side chain in the S1 specificity pocket. Apart from the five water molecules, 2101–2105, there was density interpreted as an Arg side chain but with low occupancy. Weak but still detectable density was observed at S2–S3 sites in positions analogous to the other structures (Table 4).

3.4. Disorder, effects of cryogenic freezing and radiation damage

Several side chains are disordered but most can be modelled as two alternative conformations. Of 38 Ser residues in TRY-N, seven are observed with two alternative conformations of the side chain, eight in TRY-F and six in each of the amino-acid soaked structures. Several other residues show double conformations: Thr151 in all the structures and Thr20 in all but TRY-F; Gln64 has two conformations in all the structures; Asn154 in all except TRY-N. The conserved Gln194 near the active site is only partially visible in all the structures except TRY-N. Of the charged residues, the side chain of Arg86 has two conformations in all the amino-acid soaked structures. Arg122 is invisible in TRY-N and has two conformations in the other structures. Of the hydrophobic side chains, Ile59 shows two conformers at approximately 180° to each other in all the structures. Ile77 has a second minor conformation in all the amino-acid soaked structures. Val87, Val138 and Val186 have two conformations. Pro218 has alternative sites for C^{γ} in the amino-acid soaked structures. Cys side chains are generally disordered in all the structures except TRY-N. This is accompanied by the breaking of the disulfide bonds, as described below.

In addition, all the 100 K structure but not the RT structure show varying degrees of disorder of the main chain. The regions with main-chain disorder in all or most of the cryostructures are: Gly25–Phe27, Ser59A–Tyr59C, Ser72–Gly80, Tyr94–Asn99, Arg122–Ser133, Thr144–Thr151, Val200–Thr209, Val231–Arg235. Most can be described as a movement of surface loops. In the case of residues Asp26 and Phe27 the peptide plane between them flips approximately 180°. This is accompanied by the appearance of an ordered glycerol molecule and clearly depends on the duration of soaking of the crystal in cryoprotectant solution: there is only a small indication of a change in TRY-F, which was only briefly exposed to the cryosolution, and approximately 50% conversion in the other cryostructures, which underwent extended soaking.

Comparison of the electron density of the different structures, whose exposure time ranged from less than 10 s $^{\circ -1}$ for TRY-N to 120 s $^{\circ -1}$ for TRY-LYS2 and TRY-GLN (Table 1), reveals features that change with increasing exposure time. All three disulfide bridges appear complete in TRY-N with no significant residual density in the $(F_o - F_c)$ difference map and B factors for the S atoms similar to those of the neighbouring atoms. In TRY-F two of the disulfide bridges (42–58 and 191–220) are partially broken, with significant residual peaks in the $(F_o - F_c)$ density. In the 0.81 Å structures all the disulfide bridges are partially broken. The broken bonds were modelled in alternative conformations. However, the density was difficult to interpret fully even allowing for two or three alternative conformations; it probably consists of a time average of the structure changing substantially over the course of the data collection. The occupancy factors were refined for the Cys S atoms in all the cryogenic structures. In TRY-F the apparent summed loss of sulfur density was 8%, which is probably not significant, but in TRY-ARG it was 17%, in TRY-LYS 14%, in TRY-LYS2 29% and in TRY-GLN 18%. Comparison of the cryogenic structures with TRY-N shows that the disordering of the disulfide bridges, one of which (Cys191–Cys220) closes the opening of the P1 specificity pocket, has only a localized effect on the structure.

TRY-LYS2 showed a significant deficit of $(F_o - F_c)$ density for several carboxyl and hydroxyl groups, which were well ordered in the other structures. Refinement of occupancy factors for these as well as other carbonyl and hydroxyl groups confirmed a loss of density at Ser195 (occupancy = 0.61), Ser190 (0.89), Thr159 (0.86), Asp189 (0.88), Asp201 (0.85), Asp239 (0.75) and at the C-terminal carboxyl group (0.83), while the remaining carboxyl and hydroxyl groups appeared undiminished.

4. Discussion

The resilience of *F. oxysporum* trypsin, which remains stable under only mildly inhibitory conditions and gives crystals of outstanding quality, combined with the use of synchrotron radiation allowed the detailed observation of ligands bound in the active site. Examination of the electron density of the different structures indicated a complex pattern of binding of peptide, solvent and possibly other molecules whose complete elucidation seems beyond even such an ultrahigh-resolution study. In the crystal structure, the density of the different molecular species appears to overlap as the contents of the unit cells are averaged over the crystal lattice. However, much detail can be seen at atomic (~1 Å) and at ultrahigh (higher than 1 Å) resolution. By comparing different high-resolution

Table 5
Hydrogen bonds and other close contacts between trypsin and the main-chain atoms of the ligand in the P1–P3 binding site for the TRY-* structures compared with the average values for chymotrypsin with its autolysis products (PDB code 8gch; Harel *et al.*, 1991) and the average values obtained for bovine trypsin with ten BPTI variants (Helland *et al.*, 1999).

E.s.u.s for distances and standard deviations for the mean values are given in brackets.

		Distance (Å)								
Ligand	Enzyme	TRY-N	TRY-F	TRY-ARG	TRY-LYS	TRY-LYS2	TRY-GLN	TRY-* mean	Harel *et al.* (1991) mean	Helland *et al.* (1999) mean
P3 O	216 N	3.13 (0.04)	—	3.06 (0.03)	3.05 (0.03)	3.02 (0.02)	3.06 (0.05)	3.06 (0.04)	3.16 (0.16)	3.14 (0.04)
P2 O	192 $N^{\varepsilon 2}$	3.08 (0.12)	—	—	—	—	—	—	—	2.91(0.17)
P1 N	214 O	3.05 (0.09)	—	3.07 (0.10)	3.19 (0.11)	2.87 (0.08)	3.12 (0.14)	3.06 (0.11)	3.04 (0.08)	3.35 (0.10)
	195 O^{γ}	2.99 (0.09)	—	2.90 (0.11)	2.93 (0.11)	2.99 (0.09)	2.88 (0.15)	2.94 (0.11)	3.04 (0.09)	3.01 (0.08)
P1 O†	193 N	2.92 (0.06)	2.66 (0.09)	2.63 (0.08)	2.83 (0.08)	2.82 (0.11)	2.77 (0.11)	2.77 (0.09)	2.97 (0.11)	2.66 (0.08)
	195 N	3.04 (0.06)	3.13 (0.09)	2.89 (0.08)	2.91 (0.08)	3.11 (0.11)	2.88 (0.11)	2.99 (0.09)	2.78 (0.17)	2.89 (0.06)
	195 O^{γ}	2.31 (0.06)	2.62 (0.09)	2.63 (0.08)	2.46 (0.09)	2.50 (0.11)	2.43 (0.12)	2.49 (0.09)	2.22 (0.20)	3.14 (0.07)
P1 C†	195 O^{γ}	2.06 (0.10)	1.59 (0.08)	2.04 (0.13)	2.06 (0.11)	2.24 (0.11)	1.70 (0.16)	1.95 (0.12)	2.19 (0.26)	2.71 (0.08)

† The identity of these peaks is uncertain in the TRY-* structures. The density probably includes a component of borate inhibitor, especially in TRY-N and TRY-F.

structures it was possible to identify peptides bound in the active site and after including them in the model it was possible to characterize further the remaining density.

The peptide in the binding site of TRY-N and TRY-F is probably the product of autolysis. There are seven Arg residues in the enzyme. The presence of Arg in the specificity pocket is consistent with the preference of *F. oxysporum* trypsin for Arg over Lys (Rypniewski *et al.*, 1993). The nature of the species bound at the catalytic Ser195 is uncertain. The observed density is probably the result of averaging over the crystal lattice of different ligands at low occupancy, including water, borate and enzyme-acyl reaction intermediate. The position of the peaks at the catalytic site allows an interpretation of some of them as the 'Henderson water', carbonyl O atom in the oxyanion hole or the carbonyl C atom interacting closely with the O^{γ} of the catalytic Ser195. The kinetic basis of observing reaction intermediates is uncertain (West *et al.*, 1990), but possible acyl-enzyme complexes have been reported in crystallographic studies (Harel *et al.*, 1991; Strynadka *et al.*, 1992; Yennawar *et al.*, 1994).

The density observed in sites S2 and S3 in TRY-N resembles a peptide bound in a manner similar to that found for the chymotrypsin–APPI complex (Scheidig *et al.*, 1997) or the trypsin–BPTI complex (Marquart *et al.*, 1983). The density could not be seen in TRY-F, probably as a consequence of a combination of disorder and low occupancy owing to the peptide diffusing away in the cryoprotectant solution. The density in the sites S2 and S3 could be seen again at 0.81 Å resolution, even at low occupancy, probably because of the higher resolution and intensity of the data. The P1 residue has been partly replaced in the crystals soaked in amino acids. The density in the specificity pocket is weak but significant in the 'omit maps'. Ambiguity as to the nature of the bound species can be reduced by observing the changes in the density in structures soaked with different ligands. In summary: Arg binds preferentially in the S1 pocket and its binding can be described in some detail. It is possible to displace Arg with a high concentration of Lys. Its binding in the S1 pocket is similar to that found in the bovine trypsin–BPTI complex (Marquart *et al.*, 1983). Remarkably, the density in the S2 and partly S3 sites persists at low occupancy in the different structures, but in the presence of Gln all the peptide density is diminished, in spite of the fact that Gln itself does not bind anywhere. Distinct water sites are seen in all the structures presented here, in the S1 pocket and in sites S2 and S3. They tend to occupy positions corresponding to the atoms of the ligand which are capable of forming hydrogen bonds. Such mimicry of ligands by solvent molecules has been described recently (Benini *et al.*, 1999), indicating that studying the structure of the solvent can be a useful guide in deducing or designing ligands.

The electron density near the catalytic Ser195 is probably polluted with borate, especially in TRY-N and TRY-F which were exposed to borate-free solution for less than a minute, compared with several minutes for the amino-acid soaked structures. Although the electron density suggests an acyl-enzyme intermediate similar to that observed for chymotrypsin (Harel *et al.*, 1991), the density near Ser195 should be interpreted with caution. A more reliable measure is the hydrogen-bond distance between P1 C^{α} and Ser214 O, which is similar to that found by Harel and coworkers for the autolysis products and shorter than the average distance for complexes of bovine trypsin with ten BPTI mutants (Helland *et al.*, 1999), giving an indication of closer binding at S1 of the natural reaction products than of the BPTI inhibitor (Table 5).

Although this study was not intended as an investigation of X-ray radiation damage, the comparison of our structures points to effects which seem to arise from radiation dose and cryogenic techniques. Some detailed effects of radiation damage on protein crystals have been reported recently (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000). Although the trend in recent years has been towards more brilliant X-ray sources and smaller crystals, consideration of the radiation dose should weigh on the initial choice of experimental conditions even in the case of cryogenic data collection. In any case, the effects of radiation damage should be monitored carefully and if possible discounted in order to discern properly the subtle effects of biological relevance.

An order-of-magnitude estimation, using approximations of the usable flux of the BW7B beamline, the crystal size and standard values for the elemental absorption coefficient (Creagh & Hubbell, 1995), leads to the conclusion that in a matter of hours the number of photons absorbed owing to sulfur approaches the number of S atoms in the crystal. If a substantial fraction of the absorption events leads to the disruption of the structure near the sulfur this could explain the observed deterioration of the electron-density maps. The breaking of the disulfide bridges and the apparent loss of sulfur density occurs in all our cryogenic structures and increases with the exposure to X-rays. Despite the high resolution of the data, the effects of radiation, such as breaking the disulfide bridges, increases the uncertainties in the model. Some carboxyl groups and some seryl hydroxyl groups also show a loss of density at high radiation doses.

After sulfur the next most susceptible element in the protein is oxygen, but with the absorption cross section an order of magnitude smaller it is affected significantly only at selected sites in TRY-LYS2. It is not clear why comparable loss of density is not observed in TRY-GLN, which received a similar dose of radiation. In TRY-LYS2, the loss of electron density for oxygen is observed at carboxyl groups in several acidic side chains and the C-terminus, but the highest loss of oxygen electron density occurs at the hydroxyl group of the catalytic Ser195. It is also observed to a smaller extent in the other 0.81 Å structures. The sensitivity of the carboxylate groups could be a consequence of the high local density of oxygen or to their potential for protonation/deprotonation, which could also explain the special sensitivity of Ser195. The special configuration of the catalytic site that facilitates the transfer of a proton between the Ser195 and the catalytic His57 could also reduce the bond order between the hydroxyl O and C^{β}, increasing its lability. The high susceptibility of the Ser195 hydroxyl group to radiation damage could be the reverse side of the coin in its role as the reactive group of the active site. A sensitivity to X-rays of active-site residues has also been reported for acetylcholinesterase (Ravelli & McSweeney, 2000).

There is also a noticeable decrease of the number of ordered water molecules in the crystal structures that received the highest dose of radiation (Table 2). The criteria for modelling water have been the same for all the structures, new water selection being based on statistically significant density level of the electron density. The decrease implies a disordering of the solvent at high radiation doses.

The authors would like to thank Garib Murshudov, Christoph Hermes and Victor Lamzin for useful discussion.

References

Benini, S., Rypniewski, W. R., Wilson, K. S., Miletti, S., Ciurli, S. & Mangani, S. (1999). *Struct. Fold. Des.* **7**, 205–216.

Blow, D. M. (1976). *Acc. Chem. Res.* **9**, 145–152.

Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.

Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.

Chen, Z., Li, Y., Mulichak, A. M., Lewis, S. D. & Shafer, J. A. (1995). *Arch. Biochem. Biophys.* **322**, 198–203.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Creagh, D. C. & Hubbell, J. H. (1995). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 189–206. Dordrecht/Boston/London: Kluwer Academic Publishers.

Czapinska, H. & Otlewski, J. (1999). *Eur. J. Biochem.* **260**, 571–595.

Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T. & Stroud, R. M. (1992). *Proteins*, **12**, 203–222.

Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. & Xuong, N. H. (1970). *Biochemistry*, **9**, 1997–2009.

French, S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–515.

Harel, M., Su, C. T., Frolow, F., Silman, I. & Sussman, J. L. (1991). *Biochemistry*, **30**, 5217–5225.

Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. & Smalas, A. O. (1999). *J. Mol. Biol.* **287**, 923–942.

Henderson, R. (1970). *J. Mol. Biol.* **54**, 341–354.

Kossiakoff, A. A. & Spencer, S. A. (1981). *Biochemistry*, **20**, 6462–6474.

Kraut, J. (1977). *Annu. Rev. Biochem.* **46**, 331–58.

Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E. & Powers, J. C. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 117–123.

Kretsinger, R. H. (1976). *Annu. Rev. Biochem.* **45**, 239–266.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Laskowski, M. Jr & Kato, I. (1980). *Annu. Rev. Biochem.* **49**, 593–626.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. & Laskowski, M. Jr (1997). *J. Mol. Biol.* **266**, 441–461.

Lu, W., Qasim, M. A., Laskowski, M. Jr & Kent, S. B. (1997). *Biochemistry*, **36**, 673–679.

Mac Sweeney, A., Birrane, G., Walsh, M. A., O'Connell, T., Malthouse, J. P. & Higgins, T. M. (2000). *Acta Cryst.* D**56**, 280–286.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.

Moult, J., Sussman, F. & James, M. N. (1985). *J. Mol. Biol.* **182**, 555–566.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Neurath, H. (1984). *Science*, **224**, 350–357.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.

Perona, J. J., Craik, C. S. & Fletterick, R. J. (1993). *Science*, **261**, 620–622.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Ravelli, R. B. & McSweeney, S. M. (2000). *Struct. Fold. Des.* **8**, 315–328.

Read, R. J. & James, M. N. (1988). *J. Mol. Biol.* **200**, 523–551.

Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partner Directory*, p. 86. Mountain View, CA: Silicon Graphics.

Rypniewski, W. R., Dambmann, C., von der Osten, C., Dauter, M. & Wilson, K. S. (1995). *Acta Cryst.* D**51**, 73–84.

Rypniewski, W. R., Hastrup, S., Betzel, C., Dauter, M., Dauter, Z., Papendorf, G., Branner, S. & Wilson, K. S. (1993). *Protein Eng.* **6**, 341–348.

Rypniewski, W. R., Mangani, S., Bruni, B., Orioli, P. L., Casati, M. & Wilson, K. S. (1995). *J. Mol. Biol.* **251**, 282–296.

Saxena, A. K., Singh, T. P., Peters, K., Fittkau, S., Visanji, M., Wilson, K. S. & Betzel, C. (1996). *Proteins*, **25**, 195–201.

Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). *Protein Sci.* **6**, 1806–1824.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Singer, P. T., Smalas, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *Science*, **259**, 669–673.
Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K. & James, M. N. (1992). *Nature (London)*, **359**, 700–705.
Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
West, J. B., Hennen, W. J., Lalonde, J. L., Bibbs, J. A., Zhong, Z., Meyer, E. F. Jr & Wong, C. H. (1990). *J. Am. Chem. Soc.* **112**, 5313–5320.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
Wright, C. S., Alden, R. A. & Kraut, J. (1969). *Nature (London)*, **221**, 235–242.
Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). *Biochemistry*, **33**, 7326–7336.